DTU Library

A Fast Kernel Based Searchlight Heuristic for Real-time fMRI

Hansen, Toke Jansen; Hansen, Lars Kai; Madsen, Kristoffer Hougaard

Publication date:
2010

Document Version
Early version, also known as pre-print

Citation (APA):
Hansen, T. J., Hansen, L. K., & Madsen, K. H. (2010). *A Fast Kernel Based Searchlight Heuristic for Real-time fMRI*. Poster session presented at 16th Annual Meeting of the Organization for Human Brain Mapping, Barcelona, Spain.

A Fast Kernel Based Searchlight Heuristic for Real-time fMRI

Toke Jansen Hansen[1]
Lars Kai Hansen[1]
Kristoffer Madsen[2]

Introduction

Multivariate methods recently gained much attention in the analysis of neuroimaging data, e.g., functional magnetic resonance imaging (fMRI). Compared to traditional univariate analysis, multivariate methods are able to account for distributed representational patterns and have been used successfully to, e.g., classify the human brain state in real-time [6], and to discover groups of voxels that are related to the applied stimulus [5, 4].

The support vector machine (SVM) is a commonly applied multivariate classification approach, which has proven useful in real-time fMRI as well as in localization studies (searchlight) due to good generalization performance and fast training. Furthermore, nonlinear kernelized algorithms have been used to decode complex cognitive states with a improved generalization performance compared to similar linear approaches [3], suggesting that complex interactions in brain activity patterns are important for distinguishing cognitive states.

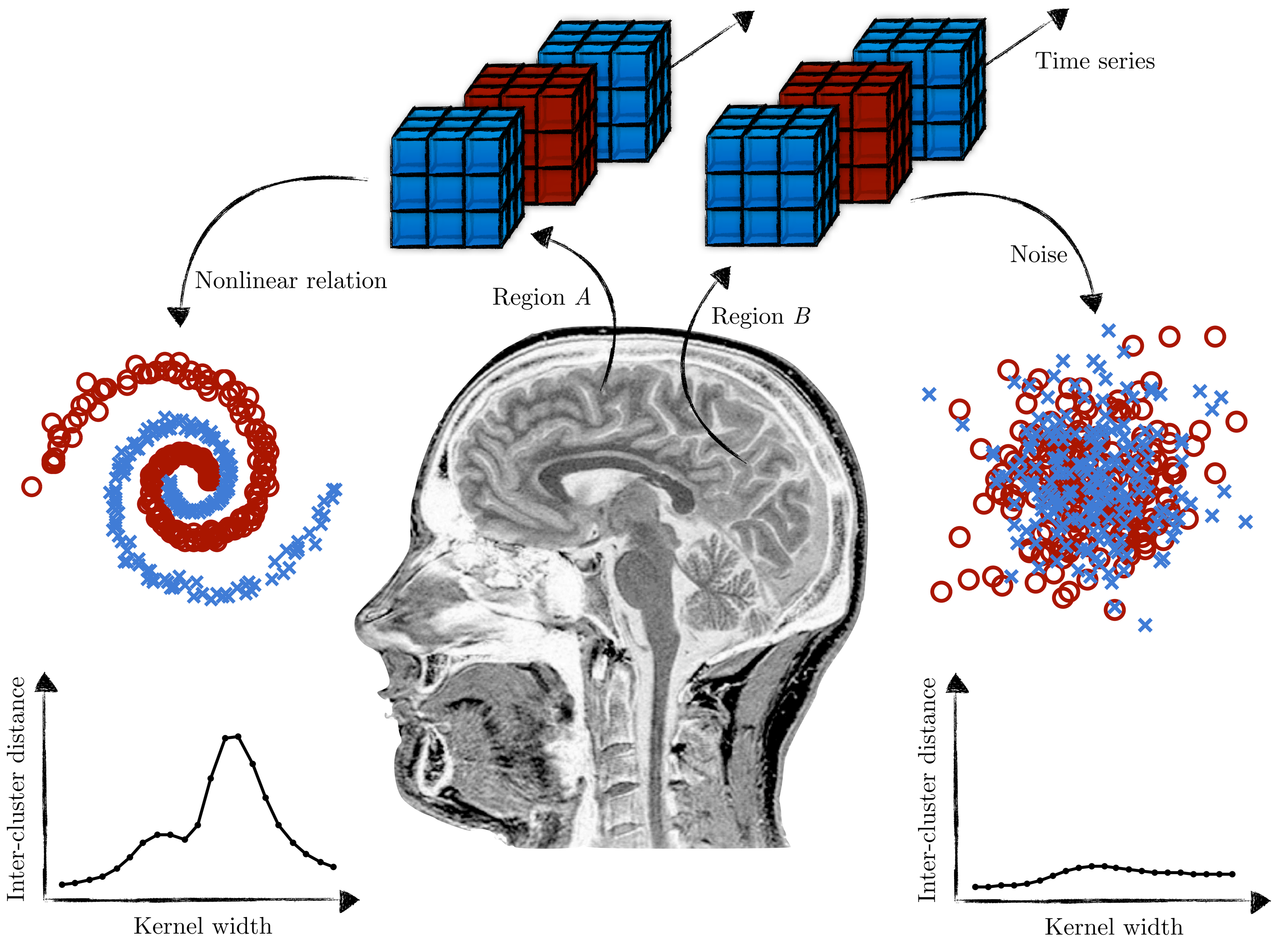

Figure 1: *Illustrates the concept of using the inter-cluster distance as a performance metric.*

In order to achieve good prediction performance most real-time fMRI studies employ a static brain mask *a priori*, and therefore depend on localizers, to account for inter-subject variability. The searchlight approach could potentially be adapted to the real-time setting, and provide a dynamic mask that takes nonlinearities within sub-regions into account. However, the computational burden of this approach can be substantial due to the resampling procedure needed to evaluate the generalization performance for each searchlight region. We here suggest a novel kernel based heuristic yielding similar results to the SVM based searchlight heuristic. The suggested approach avoids the need of a resampling procedure and hence allows computation to be carried out in a real-time setting.

Methods

An efficient heuristic for determining kernel parameters was recently proposed [8], and uses the inter-cluster distance in feature space as a quality measure for a given set of parameters. To train a kernelized SVM using the proposed method, the kernel parameters are first determined by maximizing the inter-cluster distance, after which the C parameter is determined using standard cross-validation [2].

Consider a binary classification problem, where $\mathbf{X}_+$ and $\mathbf{X}_-$ denote positive and negative samples, and where N_+ and N_- denote the number of samples in each class. In input space, $\mathcal{X}$, the distance between class means, $\bar{\mathbf{x}}_\pm$, can be computed as

$$d_{\mathcal{X}}(\bar{\mathbf{x}}_+, \bar{\mathbf{x}}_-) = d_{\mathcal{X}}\left(\frac{\sum_{\mathbf{x}_i \in \mathbf{X}_+} \mathbf{x}_i}{N_+}, \frac{\sum_{\mathbf{x}_i \in \mathbf{X}_-} \mathbf{x}_i}{N_-}\right)$$

where the function $d_{\mathcal{X}} : \mathcal{X} \times \mathcal{X} \mapsto \mathbb{R}$, is some arbitrary distance measure, -for instance the L^2-norm. Now let $\Phi : \mathcal{X} \mapsto \mathcal{F}$ be the function that maps data from the input space, to some feature space, $\mathcal{F}$, and let $d_{\mathcal{F}} : \mathcal{F} \times \mathcal{F} \mapsto \mathbb{R}$ be a distance measure in this feature space. If we choose $d_{\mathcal{F}}$ to be the L^2-norm, the distance between two points can thus be written as

$$d_{\mathcal{F}}(\Phi(\mathbf{x}), \Phi(\mathbf{y}))^2 = \Phi(\mathbf{x})^\top \Phi(\mathbf{x}) + \Phi(\mathbf{y})^\top \Phi(\mathbf{y}) - 2\Phi(\mathbf{x})^\top \Phi(\mathbf{y})$$

Since the input vectors only enters as dot products in the above expression, the kernel trick may be applied [7]

$$= k(\mathbf{x}, \mathbf{x}) + k(\mathbf{y}, \mathbf{y}) - 2k(\mathbf{x}, \mathbf{y})$$

Because the distance between two point in feature space, can be calculated in terms of kernels applied to the input space data, a similar trick can be applied to compute the L^2-norm of the class means, denoted $\bar{\mathbf{f}}_\pm$, in feature space

$$\begin{aligned} d_{\mathcal{F}}(\bar{\mathbf{f}}_+, \bar{\mathbf{f}}_-)^2 &= \frac{\sum_{\substack{\mathbf{x}_i \in \mathbf{X}_+ \\ \mathbf{x}_j \in \mathbf{X}_+}} k(\mathbf{x}_i, \mathbf{x}_j)}{N_+^2} + \frac{\sum_{\substack{\mathbf{x}_i \in \mathbf{X}_- \\ \mathbf{x}_j \in \mathbf{X}_-}} k(\mathbf{x}_i, \mathbf{x}_j)}{N_-^2} - \frac{2\sum_{\substack{\mathbf{x}_i \in \mathbf{X}_+ \\ \mathbf{x}_j \in \mathbf{X}_-}} k(\mathbf{x}_i, \mathbf{x}_j)}{N_+ N_-} \\ &= \frac{1}{N_+^2}\mathbf{1}^\top \mathbf{K}^+ \mathbf{1} + \frac{1}{N_-^2}\mathbf{1}^\top \mathbf{K}^- \mathbf{1} - \frac{2}{N_+ N_-}\mathbf{1}^\top \mathbf{K}^\pm \mathbf{1} \end{aligned}$$

Our proposed method uses the inter-cluster distance as a fast and efficient heuristic for measuring the quality of each sub-region considered by the searchlight algorithm, besides, the closed form solution completely avoids the need to store the Gram matrix for individual subregions in memory.

To handle sequential data, update rules for sample $N+1$ are expressed as follows

$$\begin{aligned} \mathbf{1}^\top \mathbf{K}_{N+1}^{\mathrm{sgn}(\mathbf{x})} \mathbf{1} &= \mathbf{1}^\top \mathbf{K}_{N}^{\mathrm{sgn}(\mathbf{x})} \mathbf{1} + 2 \cdot \mathbf{1}^\top \mathbf{k}_{N+1}^{\mathrm{sgn}(\mathbf{x})} + k_{N+1} \\ \mathbf{1}^\top \mathbf{K}_{N+1}^{\pm} \mathbf{1} &= \mathbf{1}^\top \mathbf{K}_{N}^{\pm} \mathbf{1} + \mathbf{1}^\top \mathbf{k}_{N+1}^{\neg \mathrm{sgn}(\mathbf{x})} \end{aligned}$$

and bootstrapped by

$$\mathbf{1}^\top \mathbf{K}_1^+ \mathbf{1} = k_1^+, \quad \mathbf{1}^\top \mathbf{K}_1^- \mathbf{1} = k_1^-, \quad \mathbf{1}^\top \mathbf{K}_1^\pm \mathbf{1} = k_1^\pm$$

where $\mathrm{sgn}(\mathbf{x}) \in \{+, -\}$ denotes the class of the new point $\mathbf{x}$, $\mathbf{1}^\top \mathbf{K}_N^{\mathrm{sgn}(\mathbf{x})} \mathbf{1}$ is the value from the previous iteration, $\mathbf{1}^\top \mathbf{k}_{N+1}^{\mathrm{sgn}(\mathbf{x})} = \sum_{\mathbf{x}_i \in \mathbf{X}_{\mathrm{sgn}(\mathbf{x})}} k(\mathbf{x}_i, \mathbf{x})$, and $k_{N+1} = k(\mathbf{x}, \mathbf{x})$.

In terms of computational complexity the update for a new sample requires kernel evaluations with all previous samples, thus, a single update scales linearly with the number of samples in the current kernel expansion.

Results

We tested the suggested heuristic on Blood Oxygenation Level Dependent sensitive fMRI data acquired on a 3 Tesla MR scanner (Siemens Magnetom Trio). Additional sequence parameter were as follows: 25 interleaved echo planar imaging gradient echo slices, echo time 30 ms, repetition time 1470 ms, flip angle 90 degrees. During the scanning session (800 volumes) the subject was engaged in a simple motor paradigm in which the subject was asked to respond by either left or right index finger keypress when a visual cue was presented. The classification model was used to predict with which finger the selected to press the button. Pre-processing steps included: rigid body realignment, spatial smoothing (6 mm full width at half maximum isotropic Gaussian kernel) and high pass filtering (cut-off frequency 1/128 Hz).

The searchlight approach was performed on cubic sub-regions of $3 \times 3 \times 3$ voxels, using the proposed heuristic applied with a Gaussian kernel, $k(\mathbf{x}, \mathbf{x}') = e^{-\gamma\|\mathbf{x}-\mathbf{x}'\|^2}$, to accommodate for nonlinearities in input space. For comparison we used LIBSVM [1], also with a Gaussian kernel, and for both methods we performed cross-validation to determine the parameters. For the SVM we applied cross-validation to determine the model parameters $C \in \{10^{-4}, 10^{-3}, \ldots, 10^2\}$ and $\gamma \in \{2^{-10}, 2^{-9}, \ldots, 2^{10}\}$, whereas for the suggested approach we only used cross-validation to determine the γ-parameter.

Figure 2 shows the activation map obtained from running the SVM based searchlight approach with a Gaussian kernel. Activated regions was thresholded at $> 75\%$ accuracy, corresponding to 1322 activated regions. Figure 3 shows coresponding activation map based on the suggested approach, with the threshold selected to yield 1322 activated regions, allowing direct comparison with Figure 2.

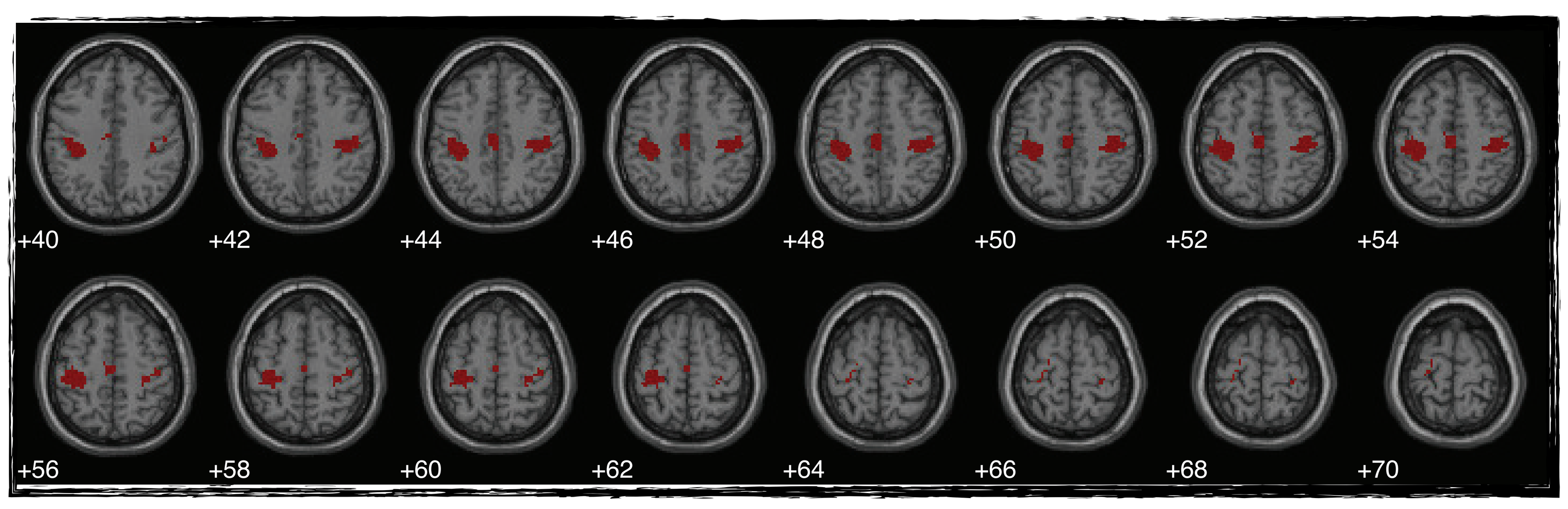

Figure 2: *Activations found using a Gaussian kernel based SVM.*

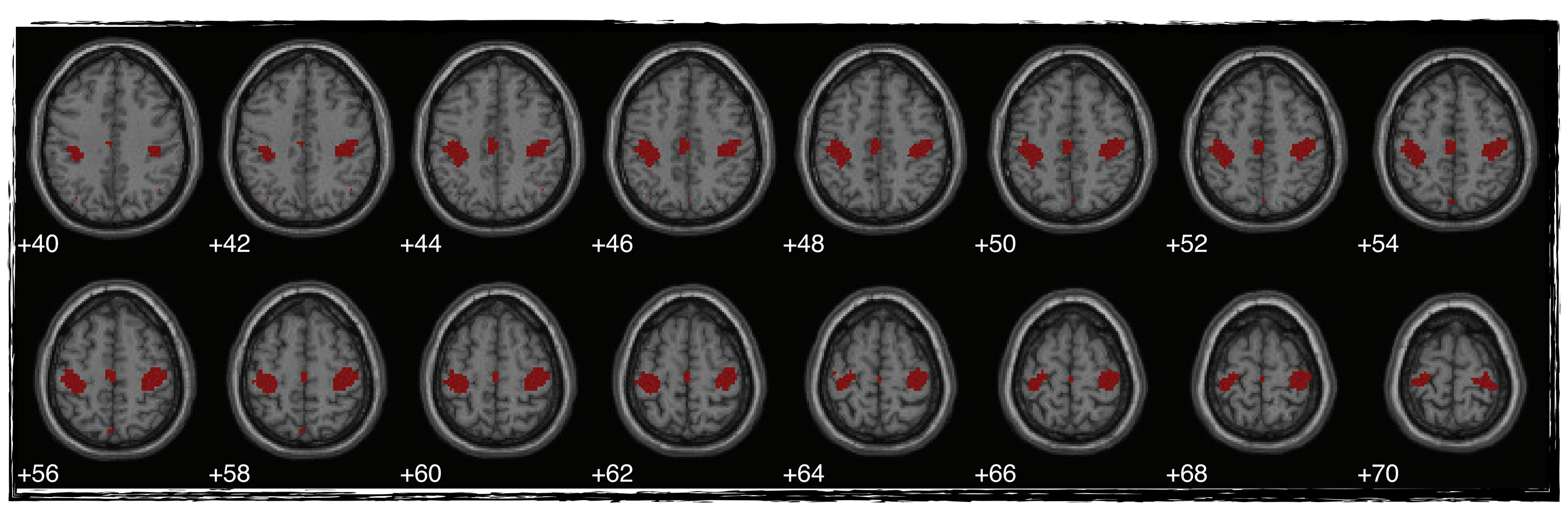

Figure 3: *Activations found using the proposed intercluster method.*

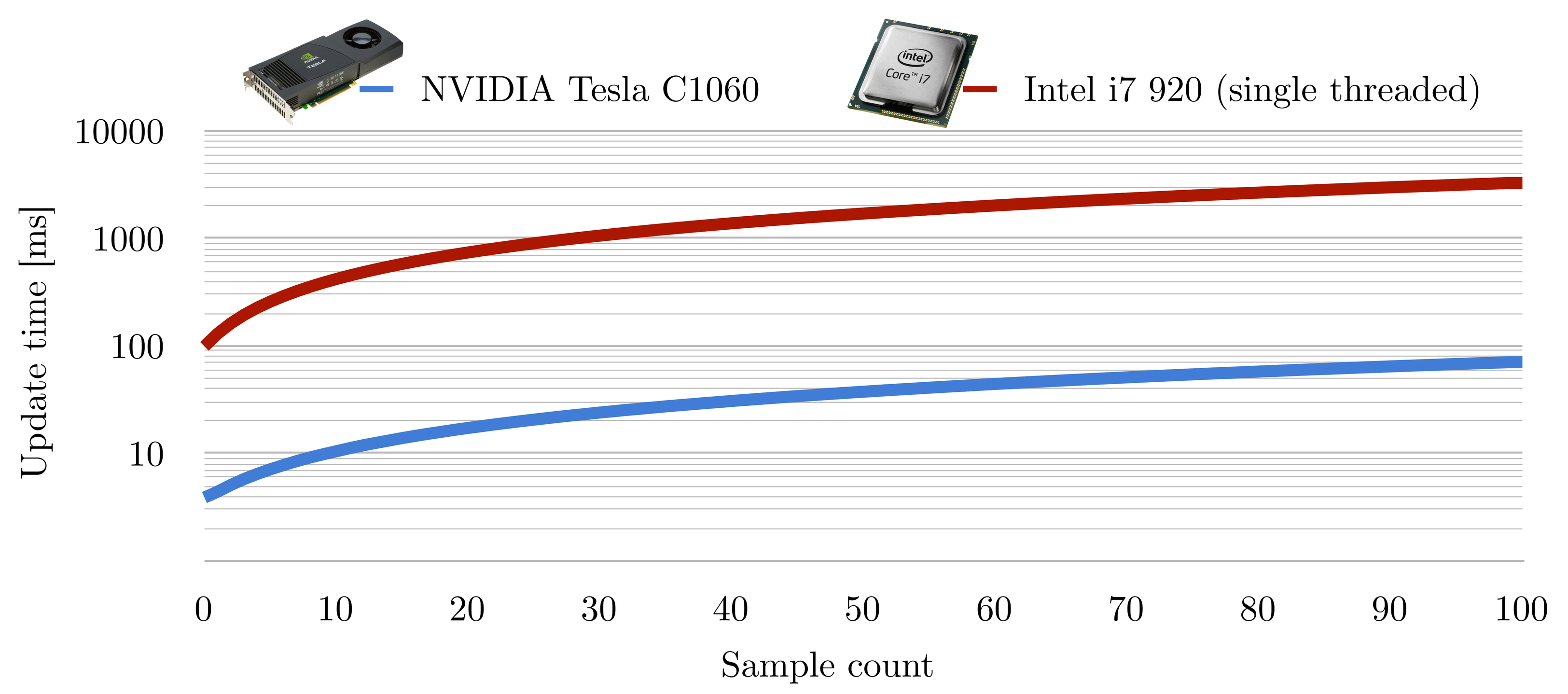

Figure 4: *Compares the performance of the heuristic when implemented for execution on respectively a GPU and CPU. The single threaded CPU implementation can handle a time window of ≈ 50 samples, before the update reaches the bound defined by the repetition time, whereas the GPU is ≈ 45 times faster.*

Conclusion

We propose a novel searchlight heuristic that yields similar results to resampling based searchlight approaches. The reduced computational complexity of the suggested heuristic enables searchlight approaches to be applied in a real-time setting preventing the need for functional localizer scans. The suggested heuristic also enables the use of dynamic searchlight procedures capable of adapting to changes in the subjects strategy, performance or brain state during the experiment. Finally, the absence of data dependencies between distinct searchlight regions and the low memory footprint, makes the heuristic highly suitable for modern multi core architectures.

References

[1] Chih-Chung Chang and Chih-Jen Lin. LIBSVM: a library for support vector machines, 2001. Software available at `http://www.csie.ntu.edu.tw/~cjlin/libsvm`.

[2] O. Chapelle, V. Vapnik, O. Bousquet, and S. Mukherjee. Choosing multiple parameters for support vector machines. Machine Learning, 46(1):131–159, 2002.

[3] C Davatzikos, K Ruparel, Y Fan, D G Shen, M Acharyya, J W Loughead, R C Gur, and D D Langleben. Classifying spatial patterns of brain activity with machine learning methods: application to lie detection. NeuroImage, 28(3):663–8, 2005.

[4] John-Dylan Haynes, Katsuyuki Sakai, Geraint Rees, Sam Gilbert, Chris Frith, and Richard E Passingham. Reading hidden intentions in the human brain. Current biology : CB, 17(4):323–8, 2007.

[5] Nikolaus Kriegeskorte, Rainer Goebel, and Peter Bandettini. Information-based functional brain mapping. Proceedings of the National Academy of Sciences of the United States of America, 103(10):3863–8, March 2006.

[6] Stephen M LaConte, Scott J Peltier, and Xiaoping P Hu. Real-time fMRI using brain-state classification. Hum Brain Mapp, 28:1033–1044, October 2007.

[7] Bernhard Scholkopf and Alexander J. Smola. Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond. MIT Press, Cambridge, MA, USA, 2001.

[8] K Wu and S Wang. Choosing the kernel parameters for support vector machines by the inter-cluster distance in the feature space. Pattern Recognition, 42(5):710–717, 2009.

DTU Informatics
Department of Informatics and Mathematical Modelling

Cognitive Systems, DTU Informatics, Denmark[1]
Danish Research Centre for Magnetic Resonance, Hvidovre Hospital, Denmark[2]

Contact Information: tjha@imm.dtu.dk